Modeling Juvenile Salmon Migration Using A Simple Markov Chain

E. Ashley STEEL, Peter GUTTORP, James J. ANDERSON, and David C. CACCIA

We describe movement patterns of hatchery-raised, juvenile, spring chinook salmon, *Oncorhynchus tshawytscha,* using a two-state Markov chain model. The existence of two states, moving and holding, is suggested by anecdotal information from a large radio telemetry study; yet adequate observations of these small-scale fish behaviors are not available for estimating transition probabilities directly. Instead, we estimate the transition probability matrix from travel times within each of 11 river segments using a method of moments approach. Bootstrapped confidence intervals are presented. Results suggest that fish behavior in the region of the confluence between the Grande Ronde and Snake Rivers includes many transitions between moving and staying while fish behavior in the Snake River is more likely to include long periods of staying.

Key Words: Inverse Gaussian distribution; Method of moments; Radio telemetry.

1. INTRODUCTION

1.1 BACKGROUND

Recent evidence indicates that substantial mortality of yearling hatchery chinook salmon, *Oncorhynchus tshawytscha,* from the Snake River system, located in Washington, Oregon, and Idaho, occurs in the free-flowing segments of the river above Lower Granite Dam (LGR); yet there is little information about fish behavior in this area. Previous research has focused on the managed sections of the river below LGR for two reasons. Historically, poor survival has been attributed to difficulties in passage through hydroelectric facilities (Raymond, 1988) and data on fish passage at hydroelectric facilities is regularly collected and readily available. However, current estimates for survival of hatchery-produced, yearling chinook salmon to LGR, the first dam encountered during seaward migration, have been as low as 15–80% and are related to distance traveled (Smith et al. 1998). These recent data suggest that an improved understanding of behavior during migration through the free-flowing segments of the river might enable improved management strategies (Independent

E. Ashley Steel is an Associate Member, National Research Center for Statistics and the Environment (NRCSE), Box 354323, University of Washington, Seattle, WA 98195 (E-mail: easteel@alumni.washington.edu). Peter Guttorp is Professor of Statistics and Director of NRCSE, University of Washington, Seattle, WA 98195. James J. Anderson is Associate Professor of Fisheries, Columbia Basin Research, Box 358218, University of Washington, Seattle, WA 98195. David C. Caccia is a Doctoral Student, NRCSE, University of Washington, Seattle, WA 98195.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 6, Number 1, Pages 80–88

Scientific Group 1996). The data set presented here provides detailed information on fish behavior in the free-flowing segments of the river above LGR reservoir.

In this paper, we develop a basic model that uses observed patterns in fish movement to estimate unobservable small-scale behavior. The model is not intended to describe actual fish movements but to provide a framework for understanding small-scale migratory behavior and for comparing this behavior between locations or over time. We incorporate mean river velocity into the basic model to approximate actual river conditions, and we examine its effect on estimated parameter values. Our aim is to develop a stochastic model that provides insights into small-scale fish behavior within the constraints of well-described, larger scale models of migration processes.

Extensions of the model introduced here might be applicable to a wide range of research efforts. Advances in radio telemetry technology have induced a proliferation of radio telemetry data; however, methods for analyzing such information are not readily available.

1.2 DATA

The fisheries data for this analysis are from a large radio telemetry study carried out by the National Marine Fisheries Service. Combination radio transmitter/passive integrated transponder (PIT) tags were surgically implanted into 129 yearling chinook salmon at Lookingglass Hatchery in March 1997. The fish were allowed to recover in the hatchery for approximately 2 weeks, after which time they were released into Lookingglass Creek. Their migration path included 132 km of the Grande Ronde River and 52 km of the Snake River. Fish behavior at the confluence of these two rivers was of particular interest.

Sample size was reduced by mortality both at the hatchery and during migration. During their migration from Lookingglass Creek to the LGR Reservoir, the fish migrated past 12 fixed-site telemetry receiving stations. Due to signal strength, antenna orientation, tag failures, and other difficulties with the electronic equipment, most fish were detected at only a subset of the 12 stations. For further details on the radio telemetry experiment, see Hockersmith, Muir, Smith, and Sanford (1998).

In this paper, we describe a technique to model travel time between stations. Travel time was calculated for each of 11 river segments for each fish observed at both endpoints. We define a river segment as the section of river between any contiguous pair of stations. There are between 7 and 31 observations for any one segment.

River velocity data was collected at 8 of the 12 observation stations during the period of out-migration. For this paper, velocity (m/sec) is defined as the maximum observable surface velocity. Velocity was estimated from the travel time of floating objects over a fixed distance and, where possible, from a boat using a Global Positioning System (GPS). Velocity for a particular river segment was indexed by the average of 2–10 measurements taken over the course of the study (April 5–May 6, 1997) at an accessible location near the telemetry receiving station. We used a single value to describe velocity over the entire course of the study because adequate data were not available at finer temporal scales. In reality, the velocity fluctuated from day to day; however, the relative velocities of the river segments should not be greatly affected by this simplification. In most river systems, velocity increases as one moves downstream; however, the final stations on the Snake River were just upstream from LGR Reservoir and velocity decreases significantly in this area.

2. MARKOV CHAIN MODEL

2.1 MODEL SUMMARY

We use the two-state Markov chain model to describe fish behavior between observation opportunities (Guttorp 1995). Our model describes a dependent, unidirectional random walk. We are interested in estimating the elements of the transition probability matrix. The transition probability matrix describes the odds of a fish being in a particular state during the next time interval given its behavior in the previous time interval. The two states included in the model are staying and moving. In each time interval, a fish either holds in the same place or moves one unit of distance downstream. The parameters of interest, p_{00}, the probability of staying given that the fish stayed in the previous time interval, and p_{11}, the probability of moving given that the fish moved in the previous time interval, define the transition probability matrix.

A two-state Markov chain model was selected to meet two criteria. First, the selected model should converge to the inverse Gaussian distribution in the limit. Previous research on migrations of large cohorts of fish between dams has shown that the distribution of travel times follows the inverse Gaussian distribution extremely well (Zabel, Anderson, and Shaw 1998). Second, the model should describe migration patterns observed in the field. During the study, mobile tracking was used to pinpoint fish locations between the fixed-site monitoring stations. Fish were often observed to stay in the same location for several days at a time before reinitiating downstream movement. A two-state Markov process is a simple model that both converges to the inverse Gaussian distribution and includes a parameter to describe periods of staying.

At regular time intervals, we assume that the fish makes a decision to move. If the fish makes a positive movement decision, then it travels some unit of distance downstream. If it makes a negative decision, then it stays in the same location. Travel time is the number of decision periods a fish must wait in order to move one unit of distance downstream. Travel time for an entire segment is estimated from the model as the time, or number of decision periods, required to make enough positive decisions to travel the length of the river segment. In the first model presented here, distance traveled per movement decision is independent of the river segment being traveled. In the second model, velocity is incorporated by allowing fish to move a distance that is dependent on the relative mean velocity in each river segment.

2.2 MODEL CALIBRATION

The model must be calibrated with two values, the number of movement decisions per hour and the distance traveled per positive movement decision. The number of movement decisions per hour translates the expected number of decisions required to make one positive decision into the expected wait time to move one unit of distance. The distance traveled with one positive movement decision defines the number of positive movement decisions required to move a given distance downstream. The combination of these two parameters defines the maximum travel speed that the model will allow. This maximum would occur if the fish were to travel at every movement opportunity. A fish may never actually travel at the maximum speed, but the value should be set so that it does not constrain the model output. We set the maximum travel speed, 32 km/h, at just over twice the maximum observed surface velocity of the river to allow for pockets of high velocity water and bursts of directional swimming.

There is a range of parameterizations by which one can achieve the appropriate maximum travel speed. For example, a fish can make one movement decision every hour and travel 32 km with a decision to move or a fish can make 1,000 decisions every hour and travel only 0.03 km with every positive movement decision. Each possible scale defines a potential model that could be used to describe fish behavior. The best scale for a particular data set is the one that provides the most information with the greatest precision. Within these constraints, it should also make biological sense.

We conducted a simulation study to assess the effect of scale on both parameter values and the width of the 95% confidence interval around the parameters. As the number of decisions per hour increased, the width of the confidence interval around $\hat{p}_{00}$ decreased while the width of the confidence interval around $\hat{p}_{11}$ increased. At 40 movement decisions per hour, confidence intervals around both estimates were small enough to provide information in both the basic and the velocity models. As well, the most information from the data, the greatest differentiation between parameter estimates in different segments of the river, was achieved at 40 movement decisions per hour for both models. At this scale, a fish travels 0.8 km with every positive decision to move. The use of this scale does not imply that 0.8 km per decision is a biologically meaningful constant with respect to individual fish behavior.

2.3 NOTATION

The following notation will be necessary for the calculations in the next section. Let

$p_{i,j}$ = probability of movement decision i during current time interval given movement decision j at previous time interval, $i, j = 0, 1$ (0 = stay, 1 = move),
w = wait time, or number of decisions, to move one unit of distance,
l_k = length (km) of river segment k, $k = 1, 2, 3, \ldots, 11$,
t_k = fish travel time (days) through river segment k,
v_k = mean water velocity (km/day) in river segment k,
$\bar{v}$ = mean of the water velocities, v_k, over all k segments,
m_k = $(l_k/0.8)/(v_k/\bar{v})$ = number of movements required to complete river segment k, and
n_k = number of fish for which travel time through segment k was observed.

The parameter v_k is equal to one for all k river segments in the basic model.

2.4 Estimation of the Transition Matrix

The first step in applying the method of moments is to calculate the expected value and the variance of t_k using the Markov model. To begin, we calculate the expectation and variance of W. We assume that each fish is initially in the move state, a reasonable assumption given that the fish must be moving to enter each study segment. The expectation and variance of W are

$$
\begin{aligned}
\mathrm{E}(W) &= \sum_{w=1}^{\infty} w p(w) \\
&= 1p_{11} + 2(1-p_{11})(1-p_{00}) + 3(1-p_{11})(1-p_{00})p_{00} \\
&\quad + 4(1-p_{11})(1-p_{00})p_{00}^2 + \cdots \\
&= p_{11} + (1-p_{11})(1-p_{00}) \sum_{x=0}^{\infty}(2+x)p_{00}^x \\
&= p_{11} + (1-p_{11})\left(2 + \frac{p_{00}}{1-p_{00}}\right)
\end{aligned}
$$

and

$$
\mathrm{var}(W) = \sum_{w=1}^{\infty} w^2 p(w) - \mathrm{E}(W)^2.
$$

$$
\begin{aligned}
\sum_{w=1}^{\infty} w^2 p(w) &= 1^2 p_{11} + 2^2(1-p_{11})(1-p_{00}) + 3^2(1-p_{11})(1-p_{00})p_{00} \\
&\quad + 4^2(1-p_{11})(1-p_{00})p_{00}^2 + \cdots \\
&= p_{11} + (1-p_{11})(1-p_{00}) \sum_{x=0}^{\infty}(2+x)^2 p_{00}^x \\
&= p_{11} + (1-p_{11})(1-p_{00}) \left[\frac{2}{(1-p_{00})^3} - \frac{1}{p_{00}} + \frac{1}{p_{00}(1-p_{00})^2}\right].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\mathrm{var}(W) &= p_{11} + (1-p_{11})(1-p_{00}) \left[\frac{2}{(1-p_{00})^3} - \frac{1}{p_{00}} + \frac{1}{p_{00}(1-p_{00})^2}\right] \\
&\quad - \left[p_{11} + (1-p_{11})\left(2 + \frac{p_{00}}{1-p_{00}}\right)\right]^2.
\end{aligned}
$$

We are interested in the moments of T_k, fish travel time through segment k. Fish travel time can be calculated as the sum of the individual wait times. These times are independent given that the initial state for each interval must be one; therefore, $\mathrm{E}(T_k) = m_k \mathrm{E}(W)$ and $\mathrm{var}(T_k) = m_k \mathrm{var}(W)$.

Because the distribution of travel times is assumed to converge to the inverse Gaussian distribution, we use the inverse Gaussian distribution to calculate the mean and variance of the data. The probability density function of a random variable, X, distributed as inverse Gaussian with parameters μ and λ is given by

$$f(x;\mu,\lambda) = \begin{cases} (\lambda/2\pi x^3)^{1/2}\exp(-\lambda(x-\mu)^2/2\mu^2 x), & x > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where μ and λ are positive (Folks and Chhikara 1978). The mean and variance of X are given by $\mathrm{E}(X) = \mu$ and $\mathrm{var}(X) = \mu^3/\lambda$. Uniform minimum variance unbiased estimates (UMVUEs) for μ and λ are $\bar{x}$ and $(n-3)/\Sigma_{i=1}^{n}(1/x_i - 1/\bar{x})$, respectively. Using UMVUEs for μ and λ, the mean of the data can be estimated by $\bar{t}_k$ and the variance by

$$\frac{1}{(n_k-3)}\cdot(\bar{t}_k)^3\cdot\sum_{n_k}\left(\frac{1}{t_k}-\frac{1}{\bar{t}_k}\right),$$

where $\bar{t}_k$ is the mean travel time for all fish observed in segment k.

The method of moments estimator of the transition probability matrix is calculated by setting the expected value and variance equal to their observed values and solving for $\hat{p}_{00}$ and $\hat{p}_{11}$. The estimates $\hat{p}_{01}$ and $\hat{p}_{10}$ can be calculated simply as $1-\hat{p}_{00}$ and $1-\hat{p}_{11}$, respectively. Although method of moments estimators are not necessarily efficient or unbiased, they are reasonable and can be obtained with a minimum of mathematical difficulty (Larsen and Marx 1986).

Confidence intervals for $\hat{p}_{00}$ and $\hat{p}_{11}$ were calculated using a nonparametric bootstrap. The travel times for each river segment were sampled with replacement. The number of randomly sampled observations was equal to the number of true observations in each segment. The estimates $\hat{p}_{00}$ and $\hat{p}_{11}$ were then calculated for each of 1,000 iterations of the sampling procedure. The 95% confidence intervals for $\hat{p}_{00}$ and $\hat{p}_{11}$ were calculated from the simulated distributions.

3. RESULTS

Tables 1 and 2 display the estimates and bootstrapped 95% confidence intervals for both p_{00} and p_{11} by river segment for the basic model (Table 1) and for the model incorporating relative velocity (Table 2). Confidence intervals were constrained to $[0,1]$. A lack of significant digits in the estimate identifies occasions where the simulated estimates were outside this range. Table 1 also includes the number of observations in each river segment and the length of that segment (km).

Velocity is incorporated into the model by dividing the reach length, l_k, by the relative mean velocity, $v_k/\bar{v}$. The effect of incorporating velocity in the model is to reduce or increase the number of movements required to complete a given river segment. We assume that, where the river is faster, a moving fish travels farther with each decision to move and vice versa. Table 2 includes the relative velocity in each segment. It contains only nine of the original segments because velocity data were not available for two of the segments.

Model fit was assessed by simulating travel times for particular river segments and comparing these simulated data to the observed travel times using a two-sample Kolmogorov–Smirnov test. There was a significant difference between the simulated and observed data when the estimates were used exactly. However, if only the estimate for p_{00} was used and the estimate for p_{11} was modified slightly, there was no significant difference.

Table 1. Estimates and Confidence Intervals for p_{00} and p_{11} From the Basic Model. Number of observations and segment length are included. GRR = Grande Ronde River, SR = Snake River.

p_{00}			River data			p_{11}		
Lower bound 95% CI	*Estimate*	*Upper bound 95% CI*	*Segment number*	n_k	L_k	*Lower bound 95% CI*	*Estimate*	*Upper bound 95% CI*
0.9773	0.9954	0.9983	GRR (2)	31	41	0.9281	0.9722	0.9834
0.7813	0.9878	0.9967	GRR (3)	11	19	0	0.8636	0.9382
0.5501	0.9970	0.9988	GRR (4)	7	11	0	0.9576	0.9701
0.8602	0.9775	0.9900	GRR (5)	16	25	0.5166	0.8792	0.9270
0.6266	0.9183	0.9714	GRR (6)	25	15	0	0.6204	0.8260
0.6608	0.9343	0.9705	GRR (7)	25	23	0	0.7253	0.8426
0.9777	0.9920	0.9955	GRR + SR (8)	21	6	0.7969	0.8756	0.9030
0.9697	0.9937	0.9973	SR (9)	23	13	0.7750	0.9117	0.9424
0.9991	0.9998	0.9999	SR (10)	20	25	0.9757	0.9893	0.9930
0.9894	0.9999	1.0000	SR (11)	20	1	0.3638	0.9451	0.9714
0.9990	1.0000	1.0000	SR (12)	20	7	0.9461	0.9914	0.9956

The required modifications of $\hat{p}_{11}$ were well within the 95% confidence interval. For example, for segment 8, the estimated value for p_{11} was 0.8756; simulations using a value of 0.8723 were not significantly different from the observed data, implying that the output from our model is consistent with our data.

Differences in behavior between estimates of the two parameters exist because the parameters affect different parts of the travel time distribution. The wait time, p_{00}, has a strong influence on the length of the tail of the inverse Gaussian distribution, which is determined by the travel times of the slowest fish. With small sample sizes, the clearest information is in the tail of the distribution; therefore, we would expect the estimates of p_{00} to be more stable and to better differentiate fish behavior among the river segments than estimates of p_{11}. This pattern was reflected in the estimates for both models. The addition of velocity had a stronger effect on $\hat{p}_{11}$ than $\hat{p}_{00}$ because the change had a greater effect on the apparent mean travel time than on the variance of travel time.

Table 2. Estimates and Confidence Intervals for p_{00} and p_{11} Incorporating River Velocity. GRR = Grande Ronde River, SR = Snake River. Relative velocity is calculated as $v_k/\bar{v}$.

p_{00}			River data		p_{11}		
Lower bound 95% CI	*Estimate*	*Upper bound 95% CI*	*Segment number*	*Relative velocity*	*Lower bound 95% CI*	*Estimate*	*Upper bound 95% CI*
0.9740	0.9953	0.9982	GRR (2)	1.10	0.8380	0.9684	0.9670
0.8453	0.9878	0.9966	GRR (3)	0.90	0	0.8787	0.9076
0.7885	0.9970	0.9988	GRR (4)	1.36	0	0.9407	0.9322
0.8820	0.9773	0.9905	GRR (5)	1.08	0.1829	0.8666	0.8659
0.8260	0.9355	0.9711	GRR (7)	1.44	0	0.5825	0.6083
0.9777	0.9920	0.9955	GRR + SR (8)	0.92	0.6916	0.8860	0.8573
0.9988	0.9998	0.9999	SR (10)	0.80	0.9661	0.9914	0.9911
0.9930	0.9999	1.0000	SR (11)	0.69	0.4341	0.9616	0.9686
0.9991	1.0000	1.0000	SR (12)	0.57	0.9514	0.9951	0.9961

In both models, estimates of p_{00} and p_{11} are substantially smaller for the river segments just upstream of the confluence of the Grande Ronde and Snake Rivers (segments 6 and 7) than estimates for any other river segments. In both the basic model and the velocity model, estimates for p_{00} in the Snake River were higher than estimated for the Grande Ronde River.

4. DISCUSSION

The two-state Markov model is a simple model to describe the process of migration in juvenile salmonids within the constraints of well-studied models of fish migration at larger scales. The method described here is successful at estimating parameters of the transition probability matrix that yield information about behavior that would be difficult to observe directly. Using this method of moments approach, estimates of the transition probability matrix can be calculated from travel time distributions, frequently observed in both radio telemetry studies and in the large PIT tag studies carried out by the National Marine Fisheries Service.

The two-state Markov model is a probabilistic model to produce estimates of unobservable yet biologically meaningful parameters. For example, p_{00}, the probability of staying given that the fish stayed in the previous time interval, gives managers and biologists an index of how long a fish might hold in a particular area. The stationary probability of staying, independent of the fish action in the previous time interval, can be calculated as $(1 - p_{11})(2 - p_{11} - p_{00})$ and provides an index of the likelihood that a fish will hold in a particular region. Because the estimates of exact parameter values can be influenced by model scale and are not intended to describe the actual physical behavior of the fish, the model will be most useful for comparing fish behavior across situations, e.g., comparing fish behavior between river segments, across years, between species, or between environments with differing conditions. Differences in fish behavior between high and low flow years or between rivers with high and low juvenile survival rates are of particular interest to fisheries managers and might be described using this approach.

Model results suggest that fish behavior differs among the different parts of the river. Smaller values of $\hat{p}_{00}$ in the region just above the confluence of the Grande Ronde and Snake Rivers for both models indicate that fish behavior may be more erratic in this area, perhaps having shorter runs of staying and holding. The very high values of $\hat{p}_{00}$ in the Snake River indicate that there may be longer runs of staying in these segments, even after adjusting for mean river velocity. Mobile tracking of radio-tagged fish during the study period also documented long periods of delay for fish migrating through the Snake River (Hockersmith et al. 1998).

In this case, the use of mean water velocity to adjust the distance traveled in a given movement did not alter the interpretation of the parameters. Further refinements of this approach might better accommodate changes in flow by using water velocity during the exact time interval in which a fish passes through a particular river segment rather than mean water velocity. The methodology described here provides an initial framework for estimating the effect of other environmental conditions, e.g., temperature and water clarity on small-scale fish behavior during migration.

ACKNOWLEDGMENTS

We would like to thank Eric Hockersmith and Dr. John Williams at the National Marine Fisheries Service for providing the fish migration data and for supporting the collection of the environmental data. Drs. Rich Zabel, Hong Qian, and Dean Billheimer provided critical assistance with the development of this project. Comments from two anonymous reviewers substantially improved the document. This research was supported by the Bonneville Power Administration through the National Marine Fisheries Service and Columbia Basin Research and by the National Research Center for Statistics and the Environment, University of Washington, Seattle.

[Received October 1998. Accepted June 2000.]

REFERENCES

Folks, J. L., and Chhikara, R. S. (1978), "The Inverse Gaussian Distribution and Its Statistical Application—A Review," *Journal of the Royal Statistical Society, Series B,* 40, 263–289.

Guttorp, P. (1995), *Stochastic Modeling of Scientific Data,* Bury St. Edmunds, UK: Chapman and Hall.

Hockersmith, E., Muir, W. D., Smith, S. G., and Sanford, B. P. (1998), "Determining Areas of Loss and Delay for Juvenile Hatchery Salmonids Above Lower Granite Dam Reservoir, 1997," Report to the Department of Energy, Bonneville Power Administration, Portland, OR (contract DE-AI79-93BP10891).

Independent Scientific Group. (1996), "Return to the River: Restoration of Salmonid Fishes in the Columbia River Ecosystem," Portland, OR: Northwest Power Planning Council.

Larsen, R. J., and Marx, M. L. (1986), *An Introduction to Mathematical Statistics and its Applications,* Englewood Cliffs, NJ: Prentice Hall.

Raymond, H. L. (1988), "Effects of Hydroelectric Development and Fisheries Enhancement on Spring and Summer Chinook Salmon and Steelhead in the Columbia River Basin," *North American Journal of Fisheries Management,* 8, 1–24.

Smith, S. G., Muir, W. D., Hockersmith, E. E., Achord, S., Eppard, M. B., Reuhle, T. E., and Williams, J. G. (1998), "Survival Estimates for Passage of Juvenile Salmonids Through Snake River Dams and Reservoirs, 1996," Report to the Department of Energy, Bonneville Power Administration, Portland, OR (contract DE-AI79-93BP108).

Zabel, R. W., Anderson, J. J., and Shaw, P. (1998), "A Multiple Reach Model to Describe the Migratory Behavior of Snake River Yearling Chinook Salmon (*Oncorhynchus tshawytscha*)," *Canadian Journal of Fisheries and Aquatic Sciences,* 55, 658–667.